JOURNAL OF CLINICAL MICROBIOLOGY, Sept. 1990, p. 1903–1905
0095-1137/90/091903-03$02.00/0

Vol. 28, No. 9

# Reproducibility and Indices of Discriminatory Power of Microbial Typing Methods

PAUL R. HUNTER

*Public Health Laboratory, City Hospital, Hoole Lane, Chester CH2 3EG, United Kingdom*

Received 12 February 1990/Accepted 15 June 1990

**When microbial strain-typing methods are compared, the most important characteristics are typeability, reproducibility, and discriminatory power. While typeability and reproducibility can be presented as numerical values, indices of discriminatory power have only recently been described. This paper examines the relationship between reproducibility and indices of discriminatory power. In an individual typing method, an inverse relationship between reproducibility and discriminatory power appears as the number of test differences required in order to distinguish between strains is increased. A method of standardizing the discriminatory power of a typing method to a predetermined reproducibility is presented. In this way the discriminatory powers of different typing methods can be compared while being standardized for the effect of reproducibility.**

When the value of any typing method is assessed, the three main characteristics that need to be considered are its typeability, reproducibility, and discriminatory power (5). The typeability of a method is the proportion of a population of distinct strains that can be assigned a type marker by that method. The reproducibility of a typing method is the proportion of strains that are typed the same on repeat testing, preferably after a period of a few months. The discriminatory power of a method is an estimate of its ability to differentiate between two unrelated strains. Gaston and I have suggested that discriminatory power can be defined mathematically as the probability that two strains chosen at random from a population of unrelated strains will be distinguished by that typing method (5). This definition led to a numerical index of discriminatory power ($D$). This probability is given by the following equation:

$$D = 1 - \frac{1}{N(N-1)} \sum_{j=1}^{s} x_j(x_j - 1) \quad (1)$$

where $s$ is the number of types, $x_j$ is the number of population members falling into the $j$th type, and $N$ is the size of the population (5). Thus, a $D$ value of 1.0 would indicate that a typing method was able to distinguish each member of a strain population from all other members of that population. Conversely, an index of 0.0 would indicate that all members of a strain population were of an identical type. An index of 0.50 would mean that if one strain was chosen at random from a strain population, then there would be a 50% probability that the next strain chosen at random would be indistinguishable from the first.

The initial equation, which is identical to Simpson's diversity index, is applicable only to situations in which all strains can be placed into mutually exclusive groups (4). To overcome this limitation, a generalized version was developed which is given by the equation

$$D = 1 - \frac{1}{N(N-1)} \sum_{j=1}^{N} a_j \quad (2)$$

where $a_j$ is the number of strains in the population which are indistinguishable from the $j$th strain and $N$ is the number of strains in the population (4). In other words, each strain in turn is compared with all other strains in the population to determine how many other strains are indistinguishable from it to give $a_j$. A computer program to calculate the value of $D$ by equation 2 has been written (P. R. Hunter, M.D. thesis, University of Manchester, Manchester, England, 1989) and is available from me. These two indices of discriminatory power can also form the basis of automatic test selection procedures when new typing methods are developed (3).

However, discriminatory power and reproducibility still have to be presented as separate values. This is a problem with typing methods that can distinguish between strains by varying numbers of test differences, particularly biotyping and bacteriophage-typing methods (2, 4). Now, the problem with equation 2 is that we can calculate several indices of discriminatory power for a single typing method as the number of test differences required in order to distinguish between strains is increased. Similarly, there can also be several reproducibility levels for a single typing method. This paper presents a method of presenting the discriminatory power of a typing method that standardizes for the effect of reproducibility.

## MATERIALS AND METHODS

Before considering the derivation of the standardized index of discriminatory power, it is important to understand the concept of test differences. Let us consider two strains, each characterized by 25 phage reactions. If all reactions are identical, there are no test differences between the two strains. If only 24 of the phage reactions are identical, then there is one test difference between the two strains, and so on for two or more test differences. Now, a number of test differences can be determined such that if there are fewer test differences than this number between the strains, then we can say that these two strains are indistinguishable. For example, if we say that two differences are required to distinguish between strains, then if there is only one difference between two test strains we cannot say they are different.

How the differing numbers of test differences affect discriminatory power can be illustrated by considering the following hypothetical population. Let us assume a popula-

tion of four strains each characterized by 10 tests, of which only the first three are variable: strain 1, + + − − + + − − + +; strain 2, + − − − + + − − + +; strain 3, + − + − + + − − + +; and strain 4, − − + − + + − − + +. In this population, strain 1 differs from strain 2 by one test difference, from strain 3 by two test differences, and from strain 4 by three test differences. If a single test difference is required to distinguish between strains, then all four strains can be distinguished from each other and the value of $D$ can be calculated from equation 1 as $D = 1.0$. However, if the tests are not completely reproducible, we may decide to require more than one test difference between strains before we can say that they are distinct. If two test differences are required, then strains 3 and 4 but not strain 2 are distinct from strain 1, strain 4 but not strain 3 or 1 is distinct from strain 2, etc. The value of $D$ for two test differences would be 0.58. If three test differences are required, then strain 4 but not strain 2 or 3 can be distinguished from strain 1, strain 2 cannot be distinguished from any other strain, etc. ($D = 0.17$). If four test differences are required, then no strain can be distinguished from any other ($D = 0.0$).

Reproducibility has already been defined in terms of the probability that, after a strain is typed on two separate occasions, the two results are deemed to be indistinguishable. For similar reasons to those given above for indices of discriminatory power, reproducibility increases as more test differences are required in order to distinguish between strains. For example, if 100 strains were characterized by 20 tests on two occasions and the results were compared, it might be found that 75 strains gave identical results on both occasions, 20 strains differed in one test, and 5 strains differed in two tests. The reproducibility by taking a single test difference as indicating nonidentity would be 75%. If two test differences were required, then the reproducibility would be 95%, and for three test differences it would be 100%. When a new typing method is developed, it is clearly a waste of discriminatory power to require more test differences than the minimum number that gives 100% reproducibility in order to distinguish between strains. Similarly, in calculating indices of discriminatory power, it is pointless to calculate those indices corresponding to numbers of test differences greater than that which gives 100% reproducibility.

With methods that can distinguish between strains by varying numbers of differences, the discrimination of a given typing method or combination of methods declines and reproducibility increases as more test differences are required in order to distinguish between strains (4). The standardized discrimination index basically determines the discrimination index of a typing method that has a reproducibility of 95%; this is designated $D_{95}$. This reproducibility value is chosen entirely arbitrarily and could in theory be any value. In most cases, it is unlikely that the reproducibility of a method would be exactly 95% for a given dissimilarity level. In this case, $D_{95}$ can be estimated by a graphical method similar to that which is used for determining MICs for 50 and 90% of tested strains (L. H. Schmidt, Antimicrob. Newsl. 4:1–8, 1987). By this method, the reproducibilities are plotted against the discriminatory indices when various numbers of test differences are required in order to distinguish between strains (Fig. 1). $D_{95}$ can then be read off the graph as the discriminatory index corresponding to 95% reproducibility.

Alternately, and more accurately, linear regression analysis can be used (1). However, as at all times when linear regression analysis is used, it is still wise to plot the relationship on graph paper to check that the relationship between discriminatory power and reproducibility is indeed linear (1). In order to obtain a linear plot, it may be necessary to plot reproducibility against the angular transformation of the discriminatory index instead of the discriminatory index itself. The angular transformation is given by the following equation (see reference 1):

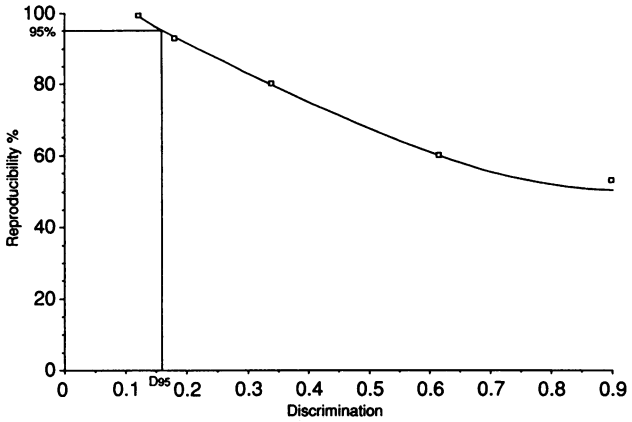


FIG. 1. Plot of reproducibility against the indices of discriminatory power for carbon source assimilation reaction typing of *C. albicans* for one to five test differences.

$$y = \sin^{-1}\sqrt{D} \quad (3)$$

That is, $y$ is the angle whose sine is $\sqrt{D}$. This usually gives an extremely good straight line over the reproducibility range of 80 to 99%; any data lying away from a straight line below this range can be safely ignored. The sine of the angle corresponding to 95% reproducibility is then squared to give the appropriate ($D_{95}$). The discriminatory powers at other reproducibility levels can be obtained as required.

## RESULTS

The combined index described above was used to analyze our published data on the comparison of various typing methods for *Candida albicans* (4). In this study, four methods (carbon source assimilation reactions, extracellular enzyme production, morphotyping, and resistotyping) were used to examine 100 unrelated strains. The reproducibility and discriminatory power for various test differences and the $D_{95}$s for these methods are shown in Table 1. $D_{95}$ was not applicable to the morphotyping code described in our previous paper (4), since only one test difference between strains was possible. The indices for some combinations of these typing methods are also shown. Table 1 also shows the application of the $D_{95}$ index to the comparison of two bacteriocin-typing methods for *Serratia marcescens*.

## DISCUSSION

The use of this single $D_{95}$ index gives a very clear impression of the relative merits of the four typing methods for *C. albicans* and reinforces our general conclusions. For a single method, resistotyping had the best combination of reproducibility and discrimination, and the best combination of methods was that of resistotyping and morphotyping. The addition of carbon source assimilation reactions to resistotyping and morphotyping was not an improvement, since although discrimination increased, reproducibility declined

TABLE 1. Reproducibilities and discriminatory powers of typing methods when differing numbers of test differences are required in order to distinguish between strains

| Species (reference) and method | Discriminatory power (% reproducibility) with indicated no. of test differences[a] | | | | | $D_{95}$ |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | |
| *C. albicans* (4) | | | | | | |
| Carbon source assimilation | 0.891 (53) | 0.607 (60) | 0.331 (80) | 0.174 (93) | 0.109 (100) | 0.155 |
| API ZYM | 0.308 (93) | 0.068 (100) | | | | 0.226 |
| Morphotype | 0.615 (89) | | | | | |
| Resistotype | 0.903 (77) | 0.694 (93) | 0.438 (97) | 0.221 (100) | | 0.564 |
| Resistotype and morphotype | 0.957 (67) | 0.814 (92) | 0.593 (97) | 0.362 (100) | | 0.673 |
| Resistotype, morphotype, and assimilation | 0.993 (40) | 0.957 (53) | 0.861 (80) | 0.718 (93) | 0.552 (100) | 0.655 |
| Resistotype, morphotype, assimilation, and API ZYM | 0.996 (40) | 0.970 (47) | 0.899 (80) | 0.774 (93) | 0.617 (100) | 0.716 |
| *S. marcescens* bacteriocin typing (6) | | | | | | |
| Spot | 0.958 (79) | 0.845 (98) | 0.662 (100) | | | 0.863 |
| Streak | 0.980 (83) | 0.911 (96) | 0.789 (100) | | | 0.919 |

[a] For definition of number of test differences required to distinguish between strains, see text.

too much. The combination of all four methods led to only a modest improvement for the additional cost and effort involved.

Similarly, for the *S. marcescens* data (Table 1), it can be seen that the index reinforces the general conclusion of Lai et al. that the cross-streaking method is more discriminatory than the spotting method (6).

The standardized index of discriminatory power described in this paper enables a much greater objectivity in the comparison of typing methods than was previously possible. However, as has already been mentioned, it has its limitations. The standardized index is applicable only to those typing methods that discriminate between strains by a number of individual tests, such as phage typing, bacteriocin typing, and biotyping. It is also potentially applicable to electrophoretic methods that present their results as the presence or absence of various bands (7, 8). The standardized index is not applicable to methods that distinguish between strains by a single test such as serotyping. Furthermore, the standardized index is not applicable to, nor is it necessary for, any typing method that has 100% reproducibility for a single test difference.

Comparing published typing methods has always been difficult and subjective. In part, this was due to the lack of an appropriate index to describe discriminatory power. Also, surprisingly, many authors do not report reproducibility studies for their methods. When published typing methods were surveyed for this paper, more than half of the papers did not mention reproducibility or described it in vague terms such as "good" or "very good." It is hoped that more diligent reporting of reproducibility, along with the use of appropriate indices of discriminatory power and the $D_{95}$ index, will greatly facilitate the assessment of new typing methods in the future.

## LITERATURE CITED


1. **Armitage, P., and G. Berry.** 1987. Statistical methods in medical research. Blackwell Scientific Publications Ltd., Oxford.
2. **Gaston, M. A.** 1987. Evaluation of a bacteriophage-typing scheme for *Enterobacter cloacae*. J. Med. Microbiol. **24**:291–295.
3. **Gaston, M. A., and P. R. Hunter.** 1989. Efficient selection of tests for bacteriological typing schemes. J. Clin. Pathol. **42**:763–766.
4. **Hunter, P. R., and C. A. M. Fraser.** 1989. Application of a numerical index of discriminatory power to a comparison of four physiochemical typing methods for *Candida albicans*. J. Clin. Microbiol. **27**:2156–2160.
5. **Hunter, P. R., and M. A. Gaston.** 1988. Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity. J. Clin. Microbiol. **26**:2465–2466.
6. **Lai, P. S., Y. F. Ngeow, S. D. Puthucheary, and C. W. Wang.** 1983. Comparison of two methods for bacteriocin typing of *Serratia marcescens*. J. Clin. Microbiol. **17**:1–6.
7. **Lee, W., J. Burnie, and R. Matthews.** 1986. Fingerprinting *C. albicans*. J. Immunol. Methods **93**:177–182.
8. **Selander, R. K., D. A. Caugant, H. Ochman, J. M. Musser, M. N. Gilmour, and T. S. Whittam.** 1986. Methods of multilocus enzyme electrophoresis for bacterial population genetics and systematics. Appl. Environ. Microbiol. **51**:873–884.